VOLUME 19, ISSUE 2, PAGES 1–23 (2018)

Criminology, Criminal Justice, Law & Society

CRIMINOLOGY, CRIMINAL JUSTICE, LAW & SOCIETY

E-ISSN 2332-886X
Available online at
https://scholasticahq.com/criminology-criminal-justice-law-society/

A Meta-Analysis of Suspect Demographic Characteristics and American Police Officer Search Decisions

P. Colin Bolger[a] and Daniel Lytle[b]

[a] *Kutztown University of Pennsylvania*
[b] *Ball State University*

A B S T R A C T A N D A R T I C L E I N F O R M A T I O N

The last summary of the research on police search decisions was completed by the National Research Council in 2004, and the present study uses the meta-analytic method to update the previous summaries of research on police search behavior. In doing so, two objectives are attempted: 1) identify which suspect characteristics impact decisions to search citizens and 2) assess the impact of methodological characteristics on research results. Of the suspect characteristics that have been analyzed, suspect race and gender appear to have the greatest impact on search decisions. However, it also appears that the methodological characteristics of research studies have a strong influence on findings. The findings of this analysis point to a primary concern for future research on search decision-making. Analyses of search behavior must standardize operationalization of the dependent variable before research in this area can be adequately synthesized.

Article History:

Received 2 October 2017
Received in revised form 31 March 2018
Accepted 4 April 2018

Keywords:

police, meta-analysis, decision-making, search

The decision for a police officer to search citizens during traffic and pedestrian stops has received a considerable amount of attention. This interest has been generated from the significant impact this decision can have on a citizen's life and the possibility that officers search certain types of people more frequently than others. Searching a citizen involves a serious deprivation of that person's liberties. The person's freedom of movement is restricted, and the state infringes upon a person's

Corresponding author: P. Colin Bolger, Department of Criminal Justice, Kutztown University of Pennsylvania, 15200 Kutztown Road, Kutztown, PA, 19530, USA.
Email: bolger@kutztown.edu

reasonable expectation of privacy in the pursuit of criminal contraband. Particularly considering the severe impacts that a search can have on a person, it is all the more troubling that there has been evidence that certain groups may be more likely to be subject to law enforcement searches than others.

Since the American Bar Foundation's "discovery" of discretion, however, it has long been recognized that legal criteria are not the only possible determinants of an officer's decision to conduct a search (Walker, 1993). In particular, many investigations have been undertaken to determine if racial minorities are more likely to be searched than White citizens (Antonovic & Knight, 2009; Engel & Calnon, 2004; Engel, Klahm, & Tillyer, 2010; Fallik & Novak, 2012; Lundman, 2004; Paoline & Terrill, 2005; Pickerill, Mosher, & Pratt, 2008; Rydberg & Terrill, 2010; Schafer, Carter, Katz-Bannister, & Wells, 2006; Tillyer, Klahm, & Engel, 2012). Police officer decisions are particularly important in this context because discovering that officers target racial minorities at higher rates may help to explain the racial disparities that exist throughout the criminal justice system. As the actors who hold the discretion to decide which people enter the system and which cases are resolved informally, police officer decisions impact the entire system. If certain types of people are treated differently by police officers, then this may be a significant contribution to the racial disparities that exist throughout the rest of the system. Generally, narrative reviews of academic research have reached this very conclusion; minority citizens are more likely to be searched than White citizens (Engel & Johnson, 2006).

The impact of suspect race, sex, and age on officer decision making has been explored in many studies. Narrative literature reviews have examined the research on these relationships (National Research Council, 2004; Riksheim & Chermak, 1993; Sherman, 1980). Most recently, the National Research Council (2004) assembled the foremost scholars to examine police research. As part of this project, these scholars examined the research around the impact of suspect race, sex, and age on police decision making. Regarding the impact of race, they found that "the evidence is mixed.… Results appear to be highly contingent on the measure of police practice, other influences that are taken into account, and the time and location context of the study" (National Research Council, 2004, p. 122-123). These scholars were not able to make a definitive decision on whether race impacts police decision making. Regarding the impact of gender, they found that "the committee is unable to draw firm conclusions about the existence of widespread bias in police practices linked to gender bias" (National Research Council, 2004, p. 121). The committee did not address the impact of suspect age in their discussion of extra-legal factors despite the presence of suspect age as a correlate in police decision making studies (see Lytle, 2014, for example).

In the 14 years since the publishing of the National Research Council's findings, no study has assessed the impact of race, gender, and age on officer search decisions across studies of police decision making. This is a fairly long drought given the previous patterns since 1980. Previously, about every 11 to 13 years, this research has been reexamined. Arrest has been assessed regarding these same correlates (see Kochel, Wilson, & Mastrofski, 2011; Lytle, 2014), and use of force has also been recently examined (see Bolger, 2015). Search decisions, however, have not been examined. Given the crucial position that search decisions pose, as they are often the first formal decision that officers make to begin the criminal justice process for citizens, it is imperative that this research be assessed so that definitive judgements may be made.

The purpose of the present study is to identify the impact of suspect characteristic on search decisions and to assess the impact of methodological characteristics of studies on research findings. To accomplish this, a meta-analysis was conducted on 17 analyses from 25 articles. This method allows for the objective assessment of the magnitude and directionality of each correlate across research studies.

Literature Review

Suspect demographic characteristics were some of the first variables examined as possible extra-legal influences on police officer decision making. A substantial amount of research effort has been placed into discerning whether who a person is impacts his or her chances of receiving different levels of police surveillance and/or coercion. Of these characteristics, race, gender, and age have received the most scholarly attention at this point. Most research that has examined the impact of a suspect's race on the likelihood of being searched has concluded that minority drivers are more likely to be searched (Engel & Calnon, 2004; Engel & Johnson, 2006; Lundman, 2004; Moon & Corley, 2007; Ridgeway, 2006; Roh & Robinson, 2009; Rojek et al., 2004; Withrow, 2004). The research on the relationship between race and likelihood of searches has not been unanimous (e.g., Tillyer and colleagues, 2012), reported that Black and Hispanic drivers were not more likely to be subjected to discretionary searches); however, this significant relationship has been found across numerous data sources and settings. For

example, Engel and Johnson (2006) reported that minority drivers were searched at higher rates by state highway police agencies. Similarly, Moon and Corley (2007) found the same results in a university setting.

Most studies on police search behavior have also included a measurement of suspect gender. Moreover, not only is the empirical evidence for this particular variable prevalent, it is almost universal in agreement. Almost every study has found that male suspects are more likely to be searched than female suspects (Engel & Calnon, 2004; Fallik & Novak, 2012; Farrell, McDevitt, Cronin, & Pierce, 2003; Lovrich, Gaffney, Mosher, Pratt, & Pickerill, 2005; Lundman, 2004; Paoline & Terrill, 2005; Pickerill et al., 2009; Rydberg & Terrill, 2010; Schafer et al., 2006; Tillyer et al., 2012). The sole contrarian is Antonovics & Knight's (2009) analysis of data from the Boston Police department, which produced a null result.

Suspect age is another well researched variable in studies of police search behavior. The predominant conclusion for studies that include this variable is that younger suspects are more likely to be searched than older suspects (Engel & Calnon, 2004; Farrell et al., 2003; Lovrich et al., 2005; Lundman, 2004; Paoline & Terrill, 2005; Pickerill et al., 2009; Rydberg & Terrill, 2010). These findings are not universal, however, as a few authors have reported mixed findings (Antonovics & Knight, 2009; Fallik & Novak, 2012; Tillyer et al., 2012). For example, Tillyer and colleagues (2012) find the negative correlation in only one of four analytic models.

The remaining variables that have been used to describe the suspect in an encounter have received substantially less attention than race, gender, and age, but some meaningful patterns can still be seen with this limited research. For suspect demeanor, the lack of interest may be due to the nonsignificant correlations produced by the few studies that have included demeanor as a predictor variable. None of the three studies that include the variable found a significant correlation between suspect demeanor and search decisions (Paoline & Terrill, 2004; Rydberg & Terrill, 2010; Tillyer et al., 2012). In contrast, the research on the effects of social class and intoxication has produced consistent evidence of the effect of these variables in opposite directions. Lower class citizens are searched more frequently than middle and upper class citizens (Engel & Calnon, 2004; Lundman, 2004; Paoline & Terrill, 2005), and suspects who show signs of intoxication are more likely to be searched (Paoline & Terrill, 2004; Rydberg & Terrill, 2010). Unfortunately, due to the limited number of studies in this area, it is not practical to meta-analyze these relationships. As a result, we restrict our analysis to race, gender, and age. However, if more studies were to become available, it would be a significant contribution to policing research to examine these relationships meta-analytically.

Prior Narrative Reviews

Three major works have attempted to draw conclusions regarding the influence of suspect race and other variables (National Research Council, 2004; Riksheim & Chermak, 1993; Sherman, 1980). These narrative literature reviews have identified four major categories of variables that may impact police officer decisions: encounter, suspect, officer, and community characteristics. Collectively, they have provided updates to the status of the research on police search decision-making; however, the most recent review was conducted over a decade ago. The preceding literature review contained an updated narrative review, and the following meta-analysis attempts to improve upon previous methods of research synthesis. Further, the National Research Council's examination could not definitively claim whether suspect demographics played a role in officer search decisions (National Research Council 2004). The National Research Council found conflicting evidence with regard to the influence of race, gender, and age on search decisions. This most recent review that was conducted by the foremost experts in the field of policing research failed to come to a definitive conclusion regarding the impacts of these potentially discriminatory relationships.

A meta-analysis represents an improvement over the narrative review approach by improving upon three methodological weaknesses. First, narrative literature reviews rely upon the expertise of the authors to adequately synthesize studies so that methodologically stronger studies are given greater weight, and the magnitude of the findings are considered rather than simply counting significant findings against nonsignificant findings. There is no guarantee that when reviewing the same works two different authors would arrive at the same conclusion. When presented with contradictory conclusions, consumers of academic research will be left without the ability to know which conclusion to trust. Additionally, the narrative method does not allow for replication by other researchers.

Second, the previous literature reviews have grouped all investigatory behaviors, such as stops and searches, into a single section. The required legal criteria for the two behaviors are different. An officer must be able to establish probable cause before he or she may search a citizen, but an officer needs only reasonable suspicion to stop a citizen. Combining these two different actions is not appropriate.

Third, previous literature reviews have treated all studies equally. This is problematic as certain studies are methodologically superior to others. For example, the Maryland Scale rates studies on their quality and shows that not all studies are created equally (Sherman et al., 1997). Moreover, Eck (2006) argues that when a number of small-*n* quasi-experimental case studies are taken together, they demonstrate the overall effectiveness of a particular intervention. These studies should not be completely discarded, but these small *n* case studies are not on the same level of methodological rigor as a longitudinal randomized controlled trial. Failing to account for variation in methodological quality could lead to potentially biased results and misjudgments on the part of the researcher.

One approach that improves upon these methodological weaknesses is meta-analysis. While not without its own limitations, a meta-analysis is designed to correct for the previously mentioned problems of subjectivity and study equality. Additionally, there is a growing body of literature that uses meta-analyses to examine criminal justice decision making (Bolger, 2015; Bontrager, Barrick, & Stupi, 2013; Daly & Bordt, 1995; Kochel et al., 2011; Lytle, 2014; Mitchell, 2005; Pratt, 1998; Wu & Spohn, 2009). Meta-analyses have explored the influence of defendant race (see Mitchell, 2005; Pratt, 1998), gender (see Bontrager et al., 2013; Daly & Bordt, 1995), and age on sentencing (Wu & Spohn, 2009). These same issues have been explored at the police stage as well. Bolger (2015) examined correlates of police use of force, Kochel and colleagues (2011) examined the influence of race on arrest, and Lytle (2014) examined the influence of race, gender, age, and ethnicity on the decision to arrest.

Method

The present study uses a meta-analytic method to assess the strength of the correlations between a variety of variables and the decision by a police officer to search a suspect. This particular method was selected because it minimizes the influence of researcher bias in comparison to the narrative review methods, and it also allows studies that analyze larger samples to receive greater weight in the estimation of the influence of the various correlates.

Sample and Selection Criteria

A search of the literature was conducted for every analysis on police search decision-making between 1960 and 2017. This lower bound of the search date range is due to the development of community policing and the general historic developments seen in American policing as a result of the civil rights movement. To ensure that every possible analysis was included, the literature search was conducted in multiple phases. First, multiple online databases were searched using the following key words: (police OR law enforcement OR officer) AND (search). The databases searched included Academic Search Complete, Academic Search Premier, Criminal Justice Abstracts, Proquest Criminal Justice, Proquest Dissertation and Theses, Proquest Research Library, Psychology and Behavioral Sciences Collection, Psyc INFO, Psyc INFO Historic, Soc Index, and Social Sciences Citation Index. The selection of these particular databases follows previous meta-analyses on criminal justice actor decision-making (Bolger, 2015; Kochel et al. 2011; Lytle, 2014; Mitchell, 2005)

These initial search results were reduced by removing duplicate studies and eliminating articles with abstracts that clearly did not describe a study of police search behavior. Articles were removed due to any of the following criteria: 1) the article described a case study, 2) the article did not provide quantitative data, 3) the article described a study on subjects who were not police officers, and/or 4) the article described a study that used an outcome measure other than search.

Second, the reference lists of the narrative literature reviews of search decision making (National Research Council, 2004; Sherman, 1980; Riksheim & Chermak, 1993) were also reviewed for any possible articles that were not identified by the online database search. Furthermore, the reference lists of those studies that were included in the sample through the database search were also reviewed for the possible identification of additional studies.

Third, in an attempt to ensure that all unpublished research on search decision making was included, recent programs from the national conventions of the American Society of Criminology and the Academy of Criminal Justice Sciences were searched for authors who had presented research on search behavior. These authors were contacted with a request for any unpublished research.

Exclusion criteria. Following the collection of every possible study, some studies were filtered out of the analysis if they did not meet additional criteria. This step was taken to ensure that the included studies were as comparable as possible. The first of these criteria was that the sample under study must have only included American local police officers. Law enforcement officials in other organizations, such as military police and federal agents, serve populations and societal roles that are significantly different from local police agencies. Local police of

other countries are also under different legal restrictions that could impact search decisions.

Second, the dependent variable had to be the search decision only. Other behaviors that could fall into the broader category of detection behaviors, such as the decision to stop a citizen, were not included. A more thorough discussion of this can be found in the description of the dependent variable. Third, only multivariate tests were included in the analysis. The findings of bivariate analyses of the search decision are far too likely to be misspecified. Research on police decision making generally, and the search decision specifically, now use much more robust research methods. Fourth, and related, the study had to include at least six covariates and needed to include all primary variables of interest, race, sex, and age of the suspect. There is not a set number of covariates required. However, one benefit of using logged odds ratios as an effect size measure is that they can be used with the presence of covariates (Fleiss & Berlin, 2009). We felt that if the study had at least six control variables, then the analysis would have controlled for the majority of potential factors that could possibly confound the relationship between our key independent variables and officer search decisions. It should also be noted that only one study had fewer than 10 covariates.[1]

Fifth, sufficient statistical information must have been reported by each study to allow for an effect size to be calculated. The specific methods used for these transformations were dependent upon the specific information present, but every effort was made to transform the reported data into a useable effect size whenever possible. Still, there were instances when the reported information was not adequate. Sixth, the study must have analyzed data from the micro police-citizen encounter level. Studies that were conducted using other research methods, such as hypothetical vignettes or surveys, were judged to not be comparable. Their inclusion would have contributed to further heterogeneity.

The full list of every study that was collected through the selection criteria is included in the index. Also included in this table is the reason for an article's exclusion, if it was excluded, or the number of analyses of different datasets that a manuscript contributed to this analysis. Overall, 68 manuscripts were identified through the selection criteria, and 17 were deemed eligible for this analysis. From these 17 different manuscripts, 23 analyses of independent effect sizes were included in the final analysis.

Multiple effect sizes from one study. A final consideration regarding the inclusion of studies into this analysis regards the inclusion of multiple analyses of the same dataset. If multiple effect sizes are included from the same dataset, this violates the independence assumption of meta-analysis. In essence, if two effect sizes are included from the same dataset, this is tantamount to counting the same effect size twice. To address this issue, the "best" effect size was selected for this analysis (e.g., see Kochel et al., 2011; Lytle, 2014; Mitchell, 2005). The decision criteria for these selections were as follows: First, the effect size produced from the study with the highest methodological quality index was selected. Second, the effect size created from the larger sample size was selected. Third, the effect size with a greater inverse of variance was selected. Fourth, the effect size produced from an analysis with a greater number of control variables was selected. In a few instances, pairs of effects sizes produced from analyses of the same dataset were identical in all four of these criteria. In these instances, analyses were run using both effect sizes, but the differences between the results of these analyses were not meaningful.

Missing data. Not all studies include all of the necessary information to be included in a meta-analysis. This is an unfortunate but an all too common occurrence when trying to conduct a meta-analysis. In particular, a handful of studies that were included in the present analysis did not report standard errors. There is no uniform method for handling missing data, but options for addressing missing standard errors include excluding studies with missing standard errors or multiple imputation (Pigott, 2009). Given that our sample size for any particular analysis was 23, and five of those did not report standard errors, we felt that excluding studies would shrink the sample size and was not appropriate. Therefore, we opted to use multiple imputation to estimate the value of standard errors. It should also be noted that we attempted to contact the authors of studies that had missing standard errors to retrieve those missing values.

Dependent Variable

Previous literature reviews have included the decision to search a suspect under the broader category of detection behaviors. Any behaviors that officers engage in to detect criminal activity and establish a rationale to intervene with a citizen would be included in this category (Riksheim & Chermak, 1993). This method treats the decision to stop a citizen for questioning the same as the decision to search a person or his or her vehicle. While the two decisions typically have the same goal, to detect criminal conduct, they are conceptually distinct due to a greater amount of coercion used against a citizen in a search. Since the officer must be able to justify a search under a greater amount of scrutiny should his or her decision be questioned, it is possible that legal factors may play a larger role in the decision to

search than the decision to stop a citizen. If these behaviors are not kept separate, then the analysis would wander dangerously close to the apples and oranges critique, and the criticism of combining different outcome measures would be valid. To avoid such a pitfall, only the decision to search a suspect was included in this analysis.

A second type of "search" study was also not included in the present study because it is dissimilar from the behavior of interest. Hit rate studies investigate whether certain groups are searched at higher rates than others and examine whether those searches of different groups result in different rates of successful searches. In particular, these studies have primarily focused on investigating whether citizens of different racial classifications are more likely to be found with criminal contraband. While certainly an important contribution to the policing literature, the purpose of this research is not the same, so these studies were not included.

The studies of interest for this analysis do not, however, measure search in a uniform way. Some studies measure search by differentiating between discretionary searches and mandatory searches. A mandatory search involves the search of a citizen that is mandated by agency policy. A search incident to arrest, for example, requires that an officer search a person during a lawful arrest to protect all other personnel from weapons, needles, and so forth when they come in contact with that citizen during further processing (such as during the booking process). On the other hand, a discretionary search is the search of a citizen when the officer has probable cause to conduct a search.

There have been multiple approaches to operationalizing these distinct search decisions into a measure of search behavior. One approach is to code mandatory searches as "no search" (Tillyer, 2014; Tillyer et al., 2012). Another approach is to code mandatory searches as a separate outcome (Fallik & Novak, 2012; Lovrich et al., 2005; Pickerill et al., 2008). Other studies, however, do not recognize the difference between these types of searches, and it must be assumed that mandatory searches are coded the same as discretionary searches (Antonovics & Knight, 2009; Briggs, 2009; Engel & Calnon, 2004; Farrell et al., 2003; Lundman, 2004; Paoline & Terrill, 2005; Rydberg & Terrill, 2010; Schafer et al., 2006). To attempt to learn whether variation in measurement impacts research findings, a moderating variable is included in this analysis.

Independent Variables

The predictor variables that were included in this analysis were those that were identified by previous narrative literature reviews and outlined in the literature review of this paper. These include suspect race, gender, and age. Because measures of race have traditionally been focused on Black individuals, this is also how we focused our operationalization of race. Studies needed to operationalize the race variable to either focus exclusively on Black individuals, or Black individuals needed to be reasonably included. A common variation on this is to dichotomize race as White and non-White. This was the case with two effect sizes in our analysis. We included these effect sizes because the authors made mention that Black individuals were included in this measure. The other variation was that race was operationalized as Black/Hispanic. We also included this effect size because Hispanic individuals are often treated similarly to Black individuals in the criminal justice system (Lytle, 2014). In both variations, the intent is to capture a concept that is similar to a Black versus White comparison.

Gender was measured as male and female across all effect sizes. Age had the most variation, but all of the operationalizations were tapping into the same construct. Age was operationalized primarily as a continuous variable. A number of studies coded age as a dichotomy of young and all other ages or used a series of dichotomies relative to a reference category. Finally, one study coded age as an ordinal variable. In each case, studies were coded so that positive values were associated with older individuals. This approach is similar to how Mitchell (2005) used multiple operationalizations for sentencing. In his meta-analysis of sentencing decisions, he had five different measures of sentencing, and they were coded in such a way that higher values indicated more punitive sanctions. We have applied the same principle to fit our age variable.

Moderating Variables

To identify the influence of methodological characteristics, a number of methodological variables are included in a moderating variable analysis. These variables include the presence of other variables in the analyses and the methods of operationalization of the independent and dependent variables. Understanding how methodological features of studies impact research findings informs future research about what concerns should be considered during primary research collection to allow for future research synthesis to be enhanced.

Variables were included in the analysis to determine whether characteristics describing the author of the study, qualities of the data collection and analysis, the inclusion of different control variables, and measurement of the independent and dependent variables impacted research findings. The variables that describe the author and study

characteristics include the discipline of the primary author, when the data for the study was initially collected, when the article was originally published, the sponsoring organization and funding, the publication outlet, where the study was conducted, the type of statistical analysis, and the total sample size. Methodological characteristics included the type of measurement used for the independent variable, whether the searches were of drivers and/or pedestrians, whether the searches included were discretionary and/or nondiscretionary, and measures of methodological quality. These moderating variables are consistent with other recent meta-analyses of criminal justice decision making (see Bolger, 2015; Kochel et al. 2011; Lytle, 2014; Mitchell, 2005).

Data Analysis

Inter-rater reliability. To assess the consistency of the coding used during data collection, inter-rater reliability statistics were calculated. This process was conducted in two stages. First, the accept/reject decisions were evaluated. Both coders reported perfect agreement on the subset of analyses chosen for the analysis. Second, for studies that were included in the analysis, the number of agreements was divided by the number of comparisons, which produced an agreement rate of 0.92. This level of agreement indicates that the coding procedures were valid and that there was a high amount of agreement between the interrater coder and the primary coder. This level of agreement is consistent with recent meta-analyses (e.g., see Bolger, 2015; Kochel et al., 2011; Lytle, 2014).

Effect size calculation. The particular effect size chosen for this analysis was a logged odds ratio. This measure has a number of advantages over other alternative effect size estimates (Lytle, 2014). First, no other effect size estimate is reported as frequently by search studies. Second, logged odds ratios can be interpreted in terms of significance and direction (Hanushek & Jackson, 1977). Third, the estimate can be transformed from other test statistics (Lipsey & Wilson, 2001). Fourth, the presence or absence of covariates does not impact the ability to make inferences about logged odds ratios (Fleiss & Berlin, 2009). Fifth, interpretation can be eased through transforming the logged odds ratio into an odds ratio. This estimate represents an increase in the dependent variable given an increase in the independent variable (Bolger, 2015).

Following the calculation of each logged odds ratio, each estimate was weighted to give greater influence to estimates generated from larger samples. Since estimates drawn from larger samples are assumed to be more representative of the population, the effect sizes were weighted by taking the product of the effect size and the inverse of the variance (Lipsey & Wilson, 2001).

These weighted estimates were then combined into mean effect sizes for each proposed correlate of search decisions. These estimates were derived from taking the quotient of the sum of the weighted effect sizes and the sum of the inverse variance for each weighted effect size estimate (Lipsey & Wilson, 2001). Effect sizes were calculated using random effects modeling. Differences between studies were assumed, a priori, to be because of genuine study differences (Lytle, 2014).[2] A 95% confidence interval was then created around each mean effect size to determine whether the estimate was significantly different from zero. The mean effect size is significantly different from zero at the .05 level if the confidence interval does not contain zero.

Effect size homogeneity. A potential limitation with a meta-analysis is that the studies analyzed are drawn from different populations. When the population is homogenous, the only variation of effect sizes is due to sampling error (Lipsey & Wilson, 2001). However, when the population is heterogeneous, characteristics of the studies could also contribute to variation in effect sizes. To estimate this possibility, the Q statistic was calculated. A significant Q statistic indicates that the population from which the studies were drawn is heterogeneous. In this case, Lipsey and Wilson (2001) recommend the use of random effects modeling when generating weighted mean effect sizes.

Publication bias. Since many of the mean effect sizes were produced from a relatively small number of effect sizes, it is necessary to assess the potential influence of publication bias. There are multiple methods to assess publication bias. Two that are commonly used are the fail-safe N and the trim and fill method (Lipsey & Wilson, 2001; Sutton, 2009). Recently, Peters, Sutton, Jones, Abrams, and Rushton (2007) found that the trim and fill method can produce inaccurate estimates of the publication bias when there is a great amount of between study heterogeneity. In other words, significant Q statistic values can cause the trim and fill method to be less accurate. Therefore, we estimate the influence of the publication bias with both the trim and fill method and the fail-safe N.

The trim and fill method imputes effect sizes of possible missing studies under a funnel plot of the standard errors and effect sizes estimates of the included studies and uses these imputed effect sizes to calculate an adjusted mean effect size (Duval & Tweedie, 2000). The trim and fill macros in STATA were used to run random effects models for these

analyses. The fail-safe *N* estimates the number of studies with opposite findings needed to render the significant effect size non-significant (Lipsey & Wilson, 2001). Higher values of the fail-safe *N* indicate a larger number of studies are needed to render the effect size non-significant.

Moderating variable analyses. Following the mean effect size calculations, those mean effect sizes based on the largest number of effect sizes were dissected through moderating variable analyses. In particular, these analyses were conducted to determine whether the methodological characteristics of studies impacted findings. They were limited, however, to those variables that were found to be correlated with search in the main effect size analysis and had a sufficient number of effects sizes to allow those effect sizes to adequately vary across the categories of the moderating variables.[3] Dissecting a mean effect size based on a handful of studies is not informative. Additionally, moderating variables were excluded if the effect sizes did not adequately fall into the different categories of the variable.

Results

Publication Characteristics

The final sample included 16 unique datasets on police officer search decisions that analyzed 25 independent effect sizes.[4] All of these studies were produced by authors affiliated with an academic institution, and the majority of analyses resulted from studies that were published in academic journals (68%). These investigations were also relatively recent endeavors; all of the studies were published since the turn of the century, and the majority of analyses were based upon data that had been collected during the same timeframe. The sample is more diverse, however, in terms of geographic region. Data based upon police departments located in the Northeast, Midwest, Southwest, the Pacific/Northwest, and multisite locations all constituted a sizable portion of the sample.

Table 1: Descriptive Statistics of Study Characteristics

Variable	Percent/Mean
Document Type	
Journal Article	68%
Professional Report	28%
Dissertation	4%
Year of Data Collection	
1990's	16%
2000's	84%
Year of Publication	
2000's	72%
2010's	28%
Discipline of Primary Author	
Criminal Justice	76%
Sociology	4%
Political Science	16%
Other	4%
Geographic Region	
Northeast	20%
Midwest	20%
Southwest	16%
Pacific/Northwest	16%
Mixed	16%

Not Reported	12%
Sample Size	1016388.96
Search Measure	
Dichotomy	92%
Ordinal Scale	8%
Type of Discretion of Searches	
Discretionary	16%
Nondiscretionary	16%
Mixed	40%
Not reported	28%
Methodological Assessment	
Theoretical rationale for variables	44%
Description of sample	100%
Representative sample	88%
Description of methods	100%
Adequate response rate	52%
Reliability reported for primary outcome	0%
Adequate reliability for primary outcome	0%
Systematic social observation	8%
Total Methodological Index	4.12

Sample Characteristics

The studies that were included in the analysis were measured for methodological quality, and most of the included studies ranked toward the lower end of the methodology index. Of the methodological characteristics used to create this index, an adequate description of the sample (100%), an adequate description of the methods (100%), and a representative sample (88%) were the most common. A theoretical explanation for the included variables (44%) and a reporting of an adequate response rates (52%) were also occasionally present. Only a small portion of the analyses (8%), however, were conducted using the systematic social observation method. Of the methodological designs used to capture police decision making behavior, systematic social observation is generally considered one of the most sound. Furthermore, no studies reported the reliability of the included search measure, and as such did not report whether the measure of the dependent variable was adequately reliable.

Table 2: Descriptive Statistics of Included Variables

Variable	Percent/Mean
Offense Seriousness	20%
Presence of Evidence	12%
Weapon	8%
Suspect Resistance	12%
Arrest	20%
Conflict at the Scene	4%
Number of Officers	4%
Number of Citizens	20%

Type of Police Intervention	16%
Location of the Incident	4%
Vehicle Condition	8%
Officer Suspicion	8%
Suspect Race	100%
Suspect Gender	100%
Suspect Age	100%
Suspect Demeanor	16%
Suspect Social Class	12%
Suspect Intoxication	4%
Officer Race	40%
Officer Gender	32%
Officer Experience	24%
Officer Education	12%
Community Economics	12%
Community Crime Rate	4%
Community Racial Demographics	16%

Q Statistic

The Q statistic assesses whether the effect sizes produced are from a homogeneous population. This statistic was calculated for each variable, and each analysis was found to have a heterogeneous population. This finding is not surprising given that our analyses encompassed police departments from across the United States. However, because of this result, when determining our weighted mean effect size for each relationship, we used random effects modeling.

Effect Sizes

Of the independent variables, race and gender were found to be correlated with search decisions. Specifically, minority individuals and males were 1.63 and 2.51 times[5] more likely to be searched respectively. Age of a suspect does not appear to influence officer search decisions.

Table 3: Mean Effect Sizes

Variable	*k*	Mean *ES*	95% *CI* Min	95% *CI* Max	*Q*
Suspect Race*	23	0.49	0.34	.65	735.44*
Suspect Gender*	23	0.82	0.67	.97	1249.04*
Suspect Age	23	-0.004	-0.02	0.01	5824.78*

* sig $p < 0.05$

Publication Bias

The results of the trim-and-fill analysis are presented in table 4. The results from the original analysis are consistent for suspect race and gender. The effect sizes for both race and gender remained significant under the trim and fill analysis as they were under the original analysis. In both cases, the *Q* statistic was significant indicating that random effects modeling should be used when processing results.

Table 4: Summary of Adjusted Significant Correlates using Trim and Fill Method

Variable	*k*	Mean *ES*	95% *CI* Min	95% *CI* Max	*Q*
Suspect Race*	20	0.38	0.20	0.56	550.71*
Suspect Gender*	26	0.483	0.330	0.64	1065.72*

* sig $p < 0.05$

Fail-safe *N* values are presented in Table 5. The fail-safe *N* statistic determines the number of null findings that would be needed to render the significant mean effect size non-significant. Our results indicate that it would take 60 and 137 studies to render our significant effect sizes for race and gender non-significant. While there is not an agreed upon value for what constitutes a satisfactory fail-safe *N* value, it seems unlikely that someone could find 60 or 137 unique studies that all have contrary findings.

Table 5: Fail-Safe *N*

Variable	Fail-Safe *N*
Race	60
Gender	137

Publication Bias

The results of the moderating variable analysis are presented in Tables 6 and 7. This analysis is meant to explore the data in more detail. This analysis dissects the overall mean effect size across various moderating variables such as characteristics of the study, the inclusion of particular variables, methodological quality, or the way in which the dependent variable is operationalized.

The effect sizes of these variables were dissected across those moderating variables with sufficient variation so that investigation of the differences across types of analyses would be informative. When all, or a vast majority, of studies fell within one category of a moderating variable, that moderating variable was not included in the analysis. These decision criteria resulted in a number of moderating variables being included in the analysis. The moderators that were selected were ones in which at least 20% of the sample was dispersed across categories. In other words, at least five effect sizes had to be in a category other than the modal category. The following publication characteristics were included as moderating variables: the type of document, whether the effect size was generated from published or unpublished sources, the decade that the manuscript was published, and the geographic region from which the data were collected. Methodological characteristics were also used as moderating variables: whether the searches were of drivers and/or pedestrians, whether the searches were discretionary and/or nondiscretionary, whether a theoretical framework was used, whether an adequate response rate was reported, whether the reliability of the outcome was adequately assessed by the original authors, and the total methodological score.. A number of potential control variables were also included from the list of encounter, suspect, officer, and community characteristics.

Through this analysis, it becomes apparent that the impact of suspect race on search decisions varies on a number of factors. First, the effect size was much larger in studies that were not published, but the mean effect size was significant in both published and unpublished studies. Second, there is some variation in the impact of race depending on where the study was conducted. Race was not significant in studies conducted in the Midwest or in studies where the region was not identified. The impact of race was strongest in the Southwest and the Northeast. Race was also significantly related to searches in the Pacific Northwest and in multi-region studies. Third, the only correlate that appeared to moderate the impact of race was seriousness of the offense. Studies that included a measure of seriousness produced a mean effect size that was smaller than those that did not include a measure of seriousness. However, seriousness did not mitigate the impact of race. Race remained significant despite the reduction in the effect size.

The relationship between race and search decisions may also be related to the quality of the study. Studies that had a theoretical explanation for the variables that were included produced a larger mean effect size than those that did not. Studies with

higher response rates and those that assessed the reliability of their outcome measure produced larger effect sizes on average. Also, we found that studies that were of the highest methodological quality showed the largest relationship between race and search. Finally, how search was operationalized appears to be related to effect sizes. Studies that included both vehicle and person searches produced the largest effect sizes. Additionally, studies that examined discretionary searches produced the largest effect sizes on average.

Table 6: Race Moderating Variable Analysis

Moderator	*k*	Mean *ES*	95% *CI* Min	95% *CI* Max
Document Type				
Journal	15	0.33	0.11	0.55
Professional Report	7	0.78	0.47	1.09
Thesis/Dissertation	1	0.66	-0.16	1.48
Publishing Format				
Published	17	0.34	0.16	0.53
Not Published	6	0.88	0.59	1.17
Publication Decade				
2000	16	0.66	0.48	0.84
2010	7	0.10	-0.17	0.38
Geographic Region				
Northeast	5	0.77	0.37	0.81
Midwest	5	0.08	-0.33	1.17
Southwest	4	0.77	0.34	0.48
Pacific Northwest	4	0.63	0.18	1.20
Mixed	2	0.22	-0.43	0.87
Unknown	3	0.30	-0.20	0.81
Seriousness Recorded				
No	18	0.52	0.36	0.69
Yes	5	0.39	0.07	0.71
Arrest Recorded				
No	20	0.51	0.34	0.68
Yes	3	0.38	-0.07	0.84
Bystanders Recorded				
No	18	0.45	0.27	0.63
Yes	5	0.64	0.29	0.99
Officer Race Recorded				
No	13	0.48	0.27	0.69
Yes	10	0.51	0.27	0.76
Officer Sex Recorded				
No	15	0.46	0.27	0.66
Yes	8	0.55	0.28	0.82

Officer Experience Recorded				
No	17	0.47	0.28	0.65
Yes	6	0.58	0.26	0.90
Methodological Quality				
Variable Explanation				
No	14	0.33	0.12	0.54
Yes	9	0.73	0.48	0.98
Response Rate Above 60%				
No	11	0.37	0.13	0.60
Yes	12	0.60	0.38	0.82
Outcome Reliability Adequate				
No	18	0.35	0.18	0.52
Yes	5	0.98	0.67	1.29
Methods Assessment Total Score				
3	13	0.33	0.11	0.55
4	2	0.50	-0.07	1.05
5	2	0.40	-0.14	0.94
6	6	0.85	0.54	1.16
Search Operationalization				
Search Type				
Exclusively Vehicle	11	0.33	0.09	0.57
Mixed	12	0.62	0.41	0.84
Discretion Type				
Discretionary	10	0.66	0.43	0.89
Nondiscretionary	5	0.39	0.07	0.71
Mixed	8	0.35	0.09	0.61

With regard to the relationship between sex and search decisions, there were some similarities to the results of the race moderator analysis but also some unique findings. First, the finding regarding differences in published versus unpublished studies is seen in sex moderating analysis. Unpublished studies produced a larger mean effect size than published, but the difference in magnitude of the effect size was smaller. Sex was significantly related to search across all geographic regions, but it was largest in studies that used multiple regions and the Northeast. The presence of a measure of seriousness of the offense appears to mitigate the impact of sex on search decisions. However, again, we find that the impact of sex is not eliminated when factoring in seriousness measures.

Table 7: Sex Moderating Variable Analysis

Moderator	*k*	Mean *ES*	95% *CI* Min	95% *CI* Max
Document Type				
Journal	15	0.86	0.64	1.08
Professional Report	7	0.79	0.50	1.08
Thesis/Dissertation	1	0.65	-0.12	1.42

Publishing Format				
Published	17	0.76	0.61	0.91
Not Published	6	0.95	0.72	1.18
Publication Decade				
2000	16	0.77	0.60	0.95
2010	7	0.95	0.67	1.24
Geographic Region				
Northeast	5	1.02	0.79	1.24
Midwest	5	0.81	0.58	1.05
Southwest	4	0.91	0.70	1.13
Pacific Northwest	4	0.48	0.27	0.69
Mixed	2	0.48	0.26	0.69
Unknown	3	0.77	0.51	1.04
Seriousness Recorded				
No	18	0.89	0.78	1.01
Yes	5	0.51	0.31	0.71
Arrest Recorded				
No	20	0.82	0.65	0.98
Yes	3	0.88	0.45	1.31
Bystanders Recorded				
No	18	0.76	0.60	0.91
Yes	5	1.04	0.75	1.34
Officer Race Recorded				
No	13	0.86	0.69	1.04
Yes	10	0.76	0.54	0.97
Officer Sex Recorded				
No	15	0.89	0.72	1.06
Yes	8	0.70	0.47	0.92
Officer Experience Recorded				
No	17	0.85	0.67	1.04
Yes	6	0.74	0.43	1.06
Methodological Quality				
Variable Explanation				
No	14	0.75	0.58	0.92
Yes	9	0.91	0.71	1.10
Response Rate Above 60%				
No	11	0.67	0.50	0.85
Yes	12	0.93	0.77	1.09
Outcome Reliability Adequate				
No	18	0.79	0.63	0.94

Yes	5	0.91	0.64	1.18
Methods Assessment Total Score				
3	13	0.71	0.53	0.88
4	2	0.95	0.52	1.38
5	2	1.15	0.72	1.59
6	6	0.88	0.65	1.11
Search Operationalization				
Search Type				
Exclusively Vehicle	11	0.74	0.55	0.92
Mixed	12	0.88	0.71	1.04
Discretion Type				
Discretionary	10	0.87	0.75	1.00
Nondiscretionary	5	0.56	0.37	0.75
Mixed	8	0.86	0.70	1.02

The relationship between suspect sex and search decisions may also be related to methodological quality. The relationship between sex and search decisions was larger in studies with a theoretical explanation for why variables were included, those with higher response rates, and those that considered the reliability of their outcome measure. Like the moderator analysis of race, higher methodological scores generally produced larger effect sizes on average compared to lower quality studies (a methods score of 3). Finally, the type of search conducted did not produce as much variation in the sex analysis. Additionally, studies that examined both discretionary and non-discretionary searches produced the largest mean effect size followed closely by studies that examined discretionary searches.

Limitations

This analysis only investigates the impact of main-effects on the search decisions. Researchers in other areas of police decision making have demonstrated that certain independent variables are stronger correlates of arrest decisions when interacting with other independent variables (Engel, Sobol, & Worden, 2000). Unfortunately, because of the nature of meta-analysis, we could not examine interaction effects.

A second limitation is true of all meta-analyses, and that is the garbage-in/garbage-out problem. Since this is a meta-analysis and is based off of other studies, this analysis is only as good as the studies that are considered. To account for this issue, we conducted a moderator analysis to examine variation in studies based on methodological quality. Results indicated that relationships were significant at the most rigorous studies. Despite this limitation, this work contributes to the search decision making literature by demonstrating the improvements that must be made by future primary research projects.

Discussion

Of the three variables, suspect race and gender appear to influence search decisions, while the age of a suspect appears to be of little consequence. However, these findings should be interpreted with caution because of the relatively small sample size of 23 unique effect sizes. Nonetheless, this finding that race and gender have a significant impact on search decisions are important when considered in the context of a Post-Ferguson police landscape. Police departments must be cognizant of potential bias in the decision making of their officers, whether conscious or unconscious and take steps to correct officer behavior and reduce bias and disparities in the criminal justice system.

These findings are bolstered, however, because they are generally in agreement with previous meta-analyses of criminal justice decision making. First, we found that the main effect of gender was significant and consistent with other decision points that have been examined. Men appear to be more likely to be searched, arrested, have force used against them, and to be sentenced more harshly (Bolger, 2015; Bontrager et al., 2013; Daly & Bordt, 1995; Lytle, 2014). This finding is less surprising given that men commit more crime than women.

Second, we found that age was not significantly related to search decisions. This is also consistent

with previous examinations of age on decision making (Bolger, 2015; Lytle, 2014; Wu & Spohn, 2009). While this finding is consistent with previous meta-analyses, it is somewhat surprising. Younger individuals may be less likely to know what rights they are guaranteed, and therefore, they may be more likely to consent to a search. However, that was not what we found in this study. Our findings indicated that younger individuals were not more or less likely to be searched than older individuals.

Third, we found that race was significantly related to search decisions. Our study marks the fourth meta-analysis of police decision making, and the fifth amongst criminal justice meta-analyses, that has indicated that there is potentially inequitable treatment based on race (Bolger, 2015; Kochel et al., 2011; Lytle, 2014; Mitchell, 2005). It appears that when the results of this study are placed in the context of the larger meta-analysis literature, suspect race has a pervasive influence in criminal justice decision making. This seems to be the case especially in policing. Black individuals appear to be more likely to be searched, arrested, and have force used against them based on our available evidence.

We also found that the relationship between race and search decisions may be contingent on several factors. The impact of race appears to be somewhat contingent on geographic region. Minority individuals in the Southwest appear to be more likely to be searched than White individuals. This is important because it is potential evidence that the influence of race on officer search decisions may depend on where the search is being conducted. A related issue is search success rates or hit rates. According to Engel and Calnon (2004), officers may have a perception that minority individuals are more likely to be carrying contraband and are thus searched more often because of the potential presence of contraband. However, when they examined a national survey, they found that minority individuals were not more likely to be carrying contraband. Moreover, Engel, Cherkauskas, Smith, Lytle, and Moore (2009) found that minority individuals who were searched actually produced lower hit rates across the state of Arizona. Hit rates were significantly higher for White individuals, but Black individuals were significantly more likely to be searched. While this study did not examine hit rates, the field of policing research could be improved by examining studies of hit rates meta-analytically.

Another important finding from our moderator analysis is that measures of seriousness mitigated but did not remove the relationships between race and search and sex and search. This is important because it is further evidence of possible bias and discrimination in the criminal justice system, but it also reinforces the strong influence of legal factors in comparison to extra-legal factors. If the criminal justice system was truly solely driven by concerns of public safety where the most serious offenders were the ones that were most likely to be punished, then the impact of race and gender would ideally be eliminated by the presence of measures of seriousness of the offense. However, that is not what we found in this study. Instead, the impact of race and gender on search decisions persisted. There is the possibility, however, that contextual factors similar to seriousness of the offense (i.e., an officer's identification of deception by the suspect) may further mitigate the influence of extra-legal, demographic factors (e.g., suspect race and sex), and further research on search decisions should continue to ensure that any potential contextual factors be identified and evaluated.

Finally, the type of search appears to be relevant to the impact of race and gender on search decisions. We found that studies that examined discretionary searches produced larger effect sizes. This is an unsurprising but important finding. The influence of suspect race upon officer discretion (whether that be from conscious racial prejudice, subconscious cognitive processes, or some other source) appears to only exist when officers exercise their discretion. When departmental policy or legal statute mandates officer behavior, officers appear to obey those regulations for the most part. While it is impossible to create a policy directing appropriate officer responses to every single possible type of scenario because of the many, many different situations in which officers find themselves, this finding indicates that effective direction leadership from administration and legislation can have its intended effect. There is still much work to be done, however, in eliminating demographic disparities in the American criminal justice system.

While it is important to identify the cause of the racial disparities that are present throughout the criminal justice system, effective policy responses can only be designed once the source(s) of these disparities has/have been identified. Policies that target racial disparities in police searches based upon the evidence of studies with weaker methodological designs may be ineffective simply because few search decisions may be impacted by the race of the suspect. If the true root cause of racial disparities originates earlier in the criminal justice system (or from societal causes outside of the system), then these policies may be targeting the symptom of the problem, not the cause, a failure that academics have bemoaned when discussing the failings of traditional police crime-control tactics and the promise of

community policing, problem-oriented policing, and intelligence-led policing.

Conclusion

The primary conclusion of this article is not, however, that more research is necessary. Instead, future research that continues to build upon this body of literature needs to be aware of a potential pitfall that has been identified by this analysis that will prevent adequate research synthesis in the future. The second interpretation of the presence of heterogeneity is that the combined effect sizes are based on analyses that use dissimilar measures of the dependent variable (Pratt, 2002), and there is also evidence that this may be an issue for the search literature. The significance of both suspect race and suspect gender are dependent upon the measure of search used in the analysis. When measures of search fail to account for the difference between mandatory and discretionary searches, males and minority citizens are more likely to be searched, but not when these different searches are kept separate. Again, this demonstrates that research studies that include more robust methods appear to report a lack of an impact of suspect race.

More robust measures of search make the distinction between mandatory and discretionary searches because the two situations represent two different scenarios that should be influenced by different factors. A search incident to arrest, for example, represents a nondiscretionary search that should be heavily influenced by encounter characteristics, such as the arrest variable, and organizational policies. Conversely, a consent search is presumed to be more likely prompted by lesser evidence, such as an officer's hunch. In these situations, extra-legal factors, such as suspect race, may have greater influence. As researchers continue to build upon the police officer search decision making body of literature, it is important to standardize measures of the dependent variable. If this is not resolved, the literature on search decision making will quickly find itself in the same situation as the research on use of force, where the behaviors included in a measure of force vary from one study to the next.

References

Antonovics, K., & Knight, B. (2009). A new look at racial profiling: Evidence from the Boston Police Department. *The Review of Economics & Statistics, 91*, 163–171. doi:10.1162/rest.91.1.163*

Bolger, P. C. (2015). Just following orders: A meta-analysis of the correlates of American police officer use of force decisions. *American Journal of Criminal Justice*, 1–27. doi:10.1007/s12103-014-9278-y

Bontrager, S., Barrick, K., & Stupi, E. (2013). Gender and sentencing: A meta-analysis of contemporary research. *Journal of Gender, Race & Justice*, *16*, 349–372.

Briggs, S. (2007). *People and places: An examination of searches during traffic stops* (Unpublished doctoral dissertation). University of Nebraska at Omaha*

Carroll, L., & Gonzalez, M. L. (2014). Out of place: Racial stereotypes and the ecology of frisks and searches following traffic stops. *Journal of Research in Crime & Delinquency, 51*, 559–584. doi:10.1177/0022427814523788 *

Daly, K., & Bordt, R. (1995). Sex effects and sentencing: An analysis of the statistical literature. *Justice Quarterly*, *12*, 141–175. doi:10.1080/07418829500092601

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of American Statistical Association, 95*, 89-98. doi:10.2307/2669529

Eck, J. (2006). When is a bologna sandwich better than sex? A defense of small-n case study evaluations. *Journal of Experimental Criminology*, *2*, 345–362. doi:10.1007/s11292-006-9014-9

Engel, R., & Calnon, J. (2004). Examining the influence of drivers' characteristics during traffic stops with police: Results from a national survey. *Justice Quarterly, 21*, 49–90. doi:10.1080/07418820400095741 *

Engel, R. S., Calnon, J. M., Liu, L., & Johnson, R. (2004). *Project on police-citizen contacts: Year 1 final report.* Retrieved from University of Cincinnati website: https://www.uc.edu/content/dam/uc/ccjr/docs/reports/project_reports/PApolicecitizenscontact0203.pdf *

Engel, R. S., Calnon, J. M., Tillyer, R., Johnson, R., Liu, L., & Wang, X. (2005). *Project on police-citizen contacts: Year 2 final report.* Retrieved from University of Cincinnati website:

https://www.uc.edu/content/dam/uc/ccjr/docs/reports/project_reports/PSP_Year_2_Report.pdf *

Engel, R. S., Chekauskas, J. C., & Smith, M. R. (2008). *Traffic stop data analysis study: Year 2 final report.* Department of Public Safety, Phoenix, AZ. Retrieved from Arizona Department of Public Safety website: http://www.azdps.gov/sites/default/files/media/Traffic_Stop_Data_Report_2008.pdf

Engel, R. S., Cherkauskas, J. C., Smith, M. R., Lytle, D., & Moore, K. (2009). *Traffic stop data analysis study: Year 3 final report.* Department of Public Safety, Phoenix, AZ. Retrieved from Arizona Department of Public Safety website: http://www.azdps.gov/sites/default/files/media/Traffic_Stop_Data_Report_2009.pdf

Engel, R., & Johnson, R. (2006). Toward a better understanding of racial and ethnic disparities in search and seizure rates. *Journal of Criminal Justice, 34*, 605–617. doi:10.1016/j.jcrimjus.2006.09.014 *

Engel, R., Klahm, C., & Tillyer, R. (2010). Citizens' demeanor, race, and traffic stops. In S. Rice & M. White (Eds.), *Race, ethnicity, and policing: New and essential readings* (pp. 287–308). New York, NY: New York University Press.

Engel, R., Sobol, J., & Worden, R. (2000). Further exploration of the demeanor hypothesis: The interaction effects of suspects' characteristics and demeanor on police behavior. *Justice Quarterly*, *17*, 235–258. doi:10.1080/07418820000096311

Engel, R. S., Tillyer, R., Cherkauskas, J. C., & Frank, J. (2007). *Traffic stop data analysis study: Year 1 final report*. Department of Public Safety, Phoenix, Retrieved from the Arizona Department of Public Safety website: https://www.azdps.gov/sites/default/files/media/Traffic_Stop_Data_Report_2007.pdf

Fallik, S., & Novak, K. (2012). The decision to search: Is race or ethnicity important? *Journal of Contemporary Criminal Justice, 28*, 146–165. doi:10.1177/1043986211425734 *

Farrell, A., McDevitt, J., Cronin, S., & Pierce, E. (2003). *Rhode Island traffic Stop Statistics Act final report*. Northeastern University Institute on Race and Justice. Retrieved from Northeastern University website: https://repository.library.northeastern.edu/files/neu:344633/fulltext.pdf

Fleiss, J., & Berlin, J. (2009). Effect sizes for dichotomous data. In H. Cooper, L. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 237–253). New York, NY: Russell Sage Foundation.

Hanushek, E., & Jackson, J. (1977). *Statistical methods for social scientists*. New York, NY: Academic Press.

Kochel, T.R., Wilson, D.B., & Mastrofski, S.D. (2011). Effect of suspect race on officers' arrest decisions. *Criminology*, *49*(2), 473–512. doi:j.1745-9125.2011.00230.x

Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Lovrich, N., Gaffney, M., Mosher, C., Pratt, T., & Pickerill, M. (2005). *Analysis of Traffic Stop Data Collected by the Washington State Patrol: Assessment of Racial and Ethnic Equity and Bias in Stops, Citations, and Searches Using Multivariate Quantitative and Multi-Method Qualitative Research Techniques: Project Final Report,* Olympia, WA: The Patrol.*

Lundman, R. (2004). Citizen race, ethnicity, and gender and citizen reports of vehicle searches by police and vehicle search rates: Toward a triangulated scholarly understanding. *Journal of Criminal Law & Criminology, 94*(2), 309–350.*

Lytle, D. (2014). The effects of suspect characteristics on arrest: A meta-analysis. *Journal of Criminal Justice, 42*(6), 589–597. doi:10.1016/j.jcrimjus.2014.10.001

Mitchell, O. (2005). A meta−analysis of race and sentencing research: Explaining the inconsistencies. *Journal of Quantitative Criminology, 21*(4), 439–466. doi:10.1007/s10940-005-7362-7

Moon, B., & Corley, C. (2007). Driving across campus: Assessing the impact of citizens' race and gender on police traffic enforcement actions. *Journal of Criminal Justice, 35*(1), 29–37. doi:10.1016/j.jcrimjus.2006.11.013

National Research Council. (2004). *Fairness and effectiveness in policing: The evidence*. Washington, D.C.: The National Academies Press.

Paoline, E., & Terrill, W. (2005). The impact of police culture on traffic stop searches: An analysis of attitudes and behavior. *Policing: IJPSM, 28*, 455–472. doi:10.1108/13639510510614555 *

Peters, J., Sutton, A., Jones, D., Abrams, K., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine, 26*(25), 4544–4562. doi:10.1002/sim.2889

Pickerill, J., Mosher, C., & Pratt, T. (2009). Search and seizure, racial profiling, and traffic stops: A disparate impact framework. *Law & Policy, 31*(1), 1–30. doi:10.1111/j.1467-9930.2008.00282.x *

Pigott, T.D. (2009). Handling missing data. In H. Cooper, L. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 399–416). New York, NY: Russell Sage Foundation.

Pratt, T.C. (1998). Race and sentencing: A meta-analysis of conflicting empirical research results. *Journal of Criminal Justice*, *26*(6), 513–523. doi:10.1016/S0047-2352(98)00028-2

Pratt, T. (2002). Meta-analysis and its discontents: Treatment destruction techniques revisited. *Journal of Offender Rehabilitation, 35*(1), 23–40. doi: 10.1300/j076v35n01_02

Ridgeway, G. (2006). Assessing the Effect of Race Bias in Post-traffic Stop Outcomes Using Propensity Scores. *Journal of Quantitative Criminology, 22*(1), 1-29.

Roh, S. & Robinson, M. (2009). A Geographic Approach to Racial Profiling: The Microanalysis and macroanalysis of Racial Disparity in Traffic Stops. *Police Quarterly, 12*, 137-169.

Rojek, J., Rosenfeld, R., & Decker, S. (2004). The Influence of Citizen's Race on Traffic Stops in Missouri. *Police Quarterly, 7*, 126-147.

Riksheim, E., & Chermak, S. (1993). Causes of police behavior revisited. *Journal of Criminal Justice, 21*(4), 353–382. doi:10.1016/0047-2352(93)90019-j

Rydberg, J., & Terrill, W. (2010). The effect of higher education on police behavior. *Police Quarterly, 13*(1), 92–120. doi:10.1177/1098611109357325 *

Schafer, J., Carter, D., Katz-Bannister, A., & Wells, W. (2006). Decision making in traffic stop encounters: A multivariate analysis of police behavior. *Police Quarterly, 9*(2), 184–209. doi:10.1177/1098611104264990 *

Sherman, L. (1980). Causes of police behavior: The current state of quantitative research. *Journal of Research in Crime & Delinquency, 17*(1), 69–100. doi:10.1177/002242788001700106

Sherman, L., Gottfredson, D., MacKenzie, D., Eck, J., Reuter, P., & Bushway, S. (1997). *Preventing crime: What works, what doesn't, what's promising: A report to the United States Congress*. Washington, DC: US Department of Justice, Office of Justice Programs.

Sutton, A.J. (2009). Publication Bias. In H. Cooper, L. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 399–416). New York, NY: Russell Sage Foundation.

Tillyer, R. (2014). Opening the black box of officer decision-making: An examination of race, criminal history, and discretionary searches. *Justice Quarterly, 31*(6), 961–985. doi:10.1080/07418825.2012.710646

Tillyer, R., Klahm, C., & Engel, R. (2012). The discretion to search: A multilevel examination of driver demographics and officer characteristics, *Journal of Contemporary Criminal Justice, 28*(2), 184–205. doi:10.1177/1043986211425721 *

Walker, S. (1993). *Taming the system: The control of discretion in criminal justice, 1950-1990*. New York, NY: Oxford University Press.

Withrow, B. (2004). Race-based Policing: A Descriptive Analysis of the Wichita Stop Study. *Police Practice and Research, 5*, 223-240.

Wu, J., & Spohn, C. (2009). Does an offender's age have an effect on sentence length? Ameta-analytic review. *Criminal Justice Policy Review*, *20*(4), 379–413. doi:10.1177/0887403409333047

About the Authors

Colin Bolger is an Assistant Professor in the Department of Criminal Justice at Kutztown University of Pennsylvania. His research focuses on police officer discretion and citizen satisfaction with policing services.

Daniel Lytle is an Assistant Professor in the Department of Criminal Justice & Criminology at Ball State University. His research focuses on meta-analysis of criminal justice decision making, police satisfaction, and fear of crime.

Appendix A

Studies Included in the Meta-Analysis

Article	Number of Analyses Included in the Meta-Analysis
Antonovics & Knight (2009)	1
Briggs (2007)	1
Carroll & Gonzalez (2014)	1
Engel & Calnon (2004)	1
Engel, Calnon, Liu, & Johnson (2004)	1
Engel et al. (2005)	1
Engel, Tillyer, Cherkauskas (2007)	1
Engel, Chekauskas, & Smith (2008)	1
Engel et al. (2009)	1
Fallik & Novak (2012)	3
Farrell et al. (2003)	1
Lovrich et al. (2005)	2
Lundman (2004)	1
Paoline & Terrill (2005)	1
Pickerill et al. (2009)	2
Rydberg & Terrill (2010)	1
Schafer et al. (2006)	3
Tillyer (2014)	1
Tillyer et al. (2012)	1

Appendix B

Studies Excluded from the Meta-Analysis

Article	Reason for Exclusion
Anonymous (1971)	3
Anwar & Fang (2006)	2
Bateman (2008)	3
Becker (2004)	2
Bland, Miller, & Quinton (2000a)	3
Bland, Miller, & Quinton (2000b)	3
Boorah (2001)	2
Boorah (2011)	2
Bowling & Phillips (2007)	3
Dominitz & Knowles (2007)	3
Ellis (2010)	3
Engel & Johnson (2006)	3
Engel & Tillyer (2008)	3
Farrell et al. (2004)	3
Gizzi (2011)	3
Gould & Mastrofski (2004)	2
Greenleaf, Skogan, & Lurigio (2008)	2
Havis & Best (2001)	2
Hernandez-Murillo & Knowles (2004)	2
Iowa Division of Criminal & Juvenile Justice Planning (2003)	3
Johnson (2006)	3
Kagehiro (1999)	1
Knowles, Persico, & Todd (1999)	2
Knowles, Persico, & Todd (2001)	2
Lichtenberg (2004)	1
Miller (2000)	3
Miller et al. (2000)	3
Miller, Quinton, & Bland (2002)	3
MP Authority (2004)	3
OÇonnor (2014)	3
Persico & Todd (2004)	2
Petrocelli, Piquero, & Smith (2003)	5
Qureshi (2010)	3
Reid Howie Associates, Ltd. (2001)	3
Ridgeway (2007)	3
Sanga (2014)	2

Schaub, Lyons, & Wagers (2000)	2
Smith & Hester (2010)	3
Smith & Petrocelli (2001)	2
Sollund (2006)	3
Souris (1966)	1
Smith et al. (2003)	3
Steward & Totman (2005)	3
Stone & Pettigrew (2000)	1
Texas Department of Public Safety (2011)	3
Tillyer & Klahm (2011)	2
Verniero & Zoubek (1999)	3
Warren & Tomaskovic-Devey (2009)	2
Wortley & Owusu-Bempah (2011)	1
Reason for Exclusion: 1: Research not conducted between 1960 and 2013 2: Sample under study not patrol officers 3: Search decision was not the dependent variable 4: Does not report data to compute ES (r) 5: Bivariate analysis 6: Not micro-level data	

Endnotes

[1] Analyses were performed with and without this one low number of covariates study. The results were not changed.

[2] All calculations for main effects and moderating variables were conducted using macros from David Wilson using STATA 13 and publically available at http://mason.gmu.edu/~dwilsonb/ma.html.

[3] Additionally, a moderator analysis was only performed on race and sex. Age could not be included in the moderator analysis because of the very small standard errors, which transformed into inverse variances became too large for STATA to handle, and elimination of particular cases would have resulted in shrinking the sample size to the point where it would not be effective to perform the moderator analysis.

[4] Each of the manuscripts that contributed an analysis for the present analysis is included in the reference list and is indicated with an asterisk.

[5] These values are the natural antilog of the effect size values presented in Table 3. Taking the natural antilog transforms the logged odds ratio into the more interpretable odds ratio.